



Б. А. Горелов, Б. Б. Горелов

РАЗРАБОТКА МОДЕЛИ ДАННЫХ ДЛЯ ЦЕЛЕЙ ОПЕРАТИВНОЙ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ФИНАНСОВОЙ ИНФОРМАЦИИ УНИВЕРСИТЕТА

Статья содержит краткое описание результатов разработки витрины данных (Data Mart), используемой в качестве основы для многомерного представления и OLAP-анализа финансовой информации в рамках информационно-аналитической системы мониторинга фактического состояния финансовых ресурсов МИФИ.

В настоящее время управление финансовыми и материальными ресурсами является центральной составляющей в общей системе управления университетом. Рыночные условия хозяйствования, демократизация системы образования, темпы изменения российского законодательства, многообразие видов деятельности в образовательных учреждениях, — все эти факторы требуют быстрой и гибкой реакции при управлении университетом, в том числе при управлении финансовыми ресурсами на базе развитых инструментальных средств, обеспечивающих глубокий и всесторонний анализ финансовых данных при принятии решений.

При использовании для анализа финансовой информации такого инструмента, как OLAP, основную сложность вызывает разработка модели данных, способной обеспечить многомерное представление анализируемой информации, что, по мнению Кодда, является самым понятным для корпоративных аналитиков способом анализа информации [10]. В данном случае как отдельные наборы данных, так и различные их агрегации могут изменяться в зависимости от внешних условий. Следовательно, каноническая реляционная модель данных [2, 4, 8] потребует построения своего расширения.

Подходы к использованию реляционной модели данных в качестве основы для многомерного представления данных развиваются в работах J. Gray [12] (использование реляционных операторов для создания куба данных);

M. Demarst [11] (проектирование и создание витрин данных); I. Mumick [13] (эксплуатация куба данных в рамках хранилища данных); N. Raden [14] (использование схемы звезды и снежинки); Л. В. Щавелева, И. А. Левенец [5, 6] (проектирование и создание аналитической системы на основе реляционного хранилища данных).

Тем не менее относительно не изученными остались вопросы построения модели данных для целей оперативной аналитической обработки финансовой информации, в частности применительно к задаче управления финансовыми данными российского университета. Причем эти вопросы приобрели особую актуальность после ввода в действие с 1 января 2002 года 25-й главы Налогового кодекса Российской Федерации, которая окончательно разделила системы бухгалтерского учета финансовой информации и учета для целей налогообложения.

1. Структура данных разрабатываемой витрины данных

Анализ предметной области показывает, что наименьшей неделимой хозяйственно-расчетной единицей университета с точки зрения системы управления финансовыми ресурсами является программно-целевой комплекс (ПЦК). ПЦК создаются на базе различных кафедр, подразделений, научно-исследовательских коллективов университета и имеют право самостоятельно расходовать собственные средства и формировать финансовые результаты своей деятельности (в рамках действующего законодательства и установленных внутренних правил), не являясь при этом юридическими лицами [1].

В рамках организационно-технической системы управления финансовыми и материальными ресурсами МИФИ ПЦК идентифициру-

ется **внутренним счетом (темой)**. Именно внутренние счета характеризуют учет расходов и поступлений финансовых средств по каждому ПЦК и в целом по университету. Учитывая, что центральной сущностью является внутренний счет, основной интерес вызывает динамика средств, имеющихся на теме. Факторами, определяющими динамику, могут являться: период времени, источники финансирования, статьи внутренней классификации расходов и поступлений, подразделения, статьи бюджетной классификации расходов. Следовательно, при многомерном представлении финансовой информации в ячейках куба данных [3] в роли количественных показателей (мер), характеризующих анализируемый процесс, будут выступать суммы расходов и поступлений финансовых средств. При этом проектируемый куб данных будет содержать пять измерений, которые определяют систему координат многомерного пространства финансовых данных (рис. 1):

— измерение «источники» — позволяет рассматривать суммы расходов и поступлений финансовых средств в зависимости от форм финансирования, видов деятельности, источников финансирования и внутренних счетов;

— измерение «внутренние статьи» — позволяет представить анализируемую информацию в соответствии с внутренней классификацией расходов и поступлений финансовых средств;

— измерение «бюджетные статьи» — позволяет анализировать суммы финансовых средств в соответствии с предметными статьями расходов бюджетной классификации;

— измерение «время» — позволяет рассматривать суммы расходов и поступлений финансовых средств в зависимости от даты совершения финансовой операции;

— измерение «подразделения» — позволяет анализировать финансовую информацию в зависимости от номеров структурных подразделений университета, к которым относятся суммы финансовых средств.

Поскольку для целей анализа и управления финансовыми ресурсами интерес представляют как детализированные, так и суммарные показатели, то на множестве значений каждого измерения, в соответствии со спецификой предметной области, устанавливаются отношения иерархии. Необходимо отметить, что список уровней иерархии перечисленных измерений является неполным, поскольку в рамках многомерного представления данных по умолчанию предполагается, что каждое измерение содержит нулевой уровень, на котором объединяются все значения данного измерения. Например, при использовании нулевого уровня измерения «время» производится агрегирование по всем годам, которые определены в рамках рассматриваемого измерения.

Измерение «источники» имеет четыре уровня иерархии (рис. 2). Первый уровень определяет формы финансирования (бюджетные средства, внебюджетные средства, накладные расходы, целевые средства и т. д.). Второй уровень определяет виды деятельности (научно-исследовательская, образовательная деятельность и т. п.), а также министерства и ведом-

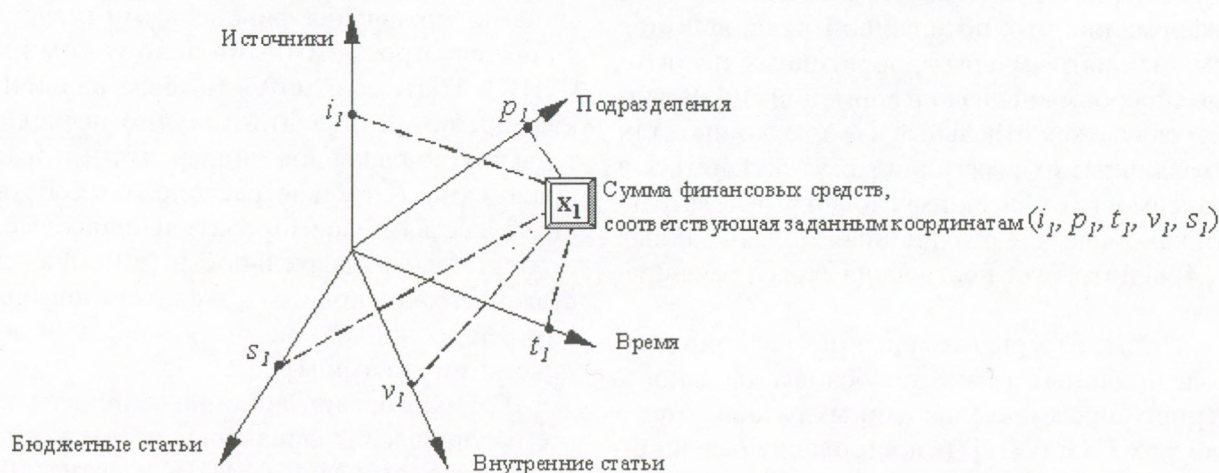


Рис. 1. Схема многомерного представления финансовых данных



ства, финансирующие университет (РФФИ, Министерство науки, Министерство образования и т. д.). Третий уровень иерархии, в соответствии со сложившейся внутренней классификацией, определяет источники финансирования (1, 5, 3, 4, ю, 8 и т. д.). Четвертый уровень определяет номера внутренних счетов (80-4-609-154, 80-8-910-011 и т. п.).

Измерение «внутренние статьи» имеет два уровня иерархии (рис. 3). Первый уровень определяет типы финансовых операций (приход, расход). Второй уровень определяет номера статей внутренней классификации расходов и поступлений финансовых средств (001, 003, 100, 120, 305, 801 и т. д.).

Измерение «бюджетные статьи» имеет шесть уровней иерархии (рис. 4). Относительно статей бюджетной классификации расходов необходимы некоторые пояснения. Все бюджетные статьи подразделяются на шесть групп. Каждый элемент вышестоящей группы по смысловой нагрузке включает в себя некоторое подмножество элементов подчиненной группы. Например, статья 110000 «за-

купки товаров и оплата услуг» содержит в себе следующие статьи: 110100 — «оплата труда государственных служащих»; 110200 — «начисления на оплату труда (страховые взносы на государственное социальное страхование граждан)» и др. В свою очередь статья 110100 включает в себя: 110110 — «оплата труда гражданских служащих»; 110140 — «оплата труда внештатных сотрудников», и т. д. Поэтому иерархия данного измерения определена в соответствии с иерархией бюджетных статей.

Измерение «время» имеет четыре уровня иерархии, которые соответствуют сложившимся отчетным периодам, — год, квартал, месяц, число (рис. 5).

Измерение «подразделения» имеет четыре уровня иерархии (рис. 6). Первый уровень определяет типы структурных подразделений университета (факультеты и части). Второй уровень иерархии определяет названия факультетов и частей (режимная часть, финансовая часть, информационная часть, а также факультеты «К», «А», «Т» и т. д.). Третий уровень оп-

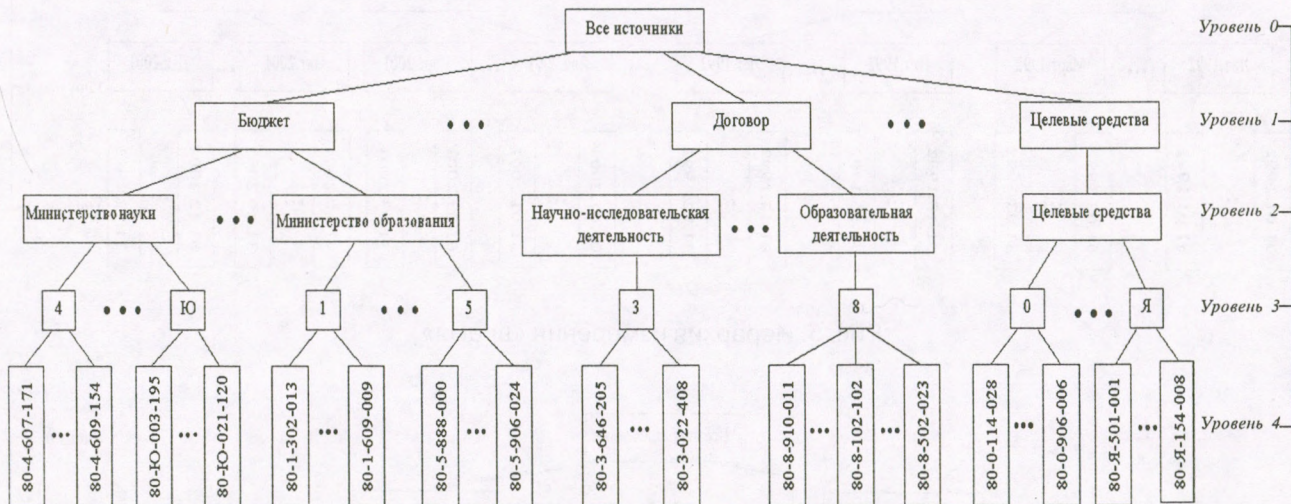


Рис. 2. Иерархия измерения «источники»



Рис. 3. Иерархия измерения «внутренние статьи»

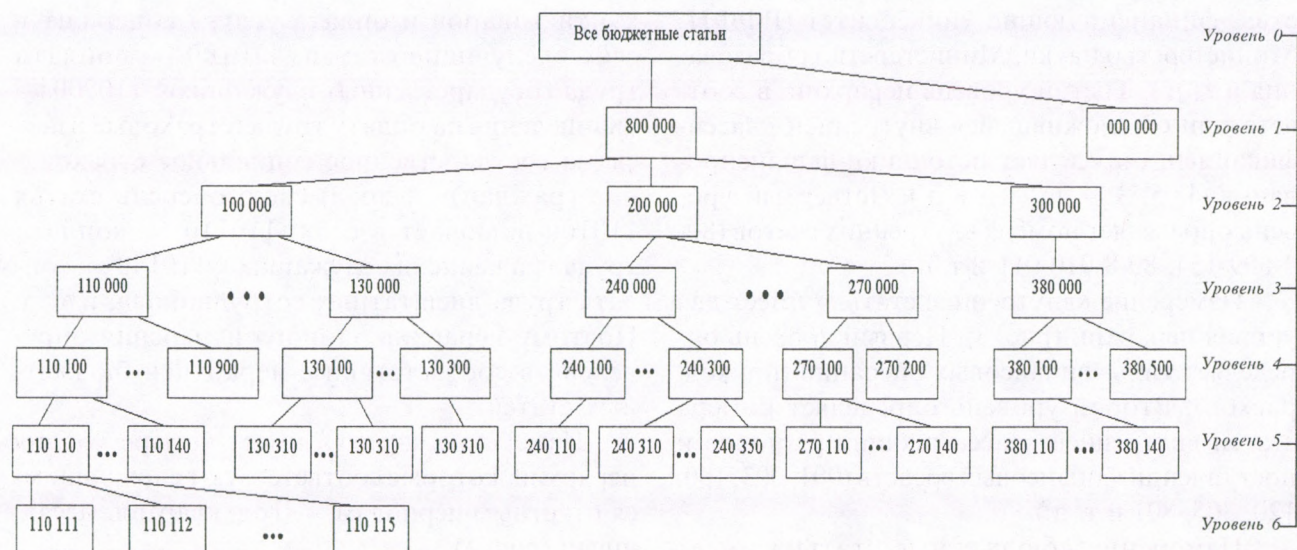


Рис. 4. Иерархия измерения «бюджетные статьи»

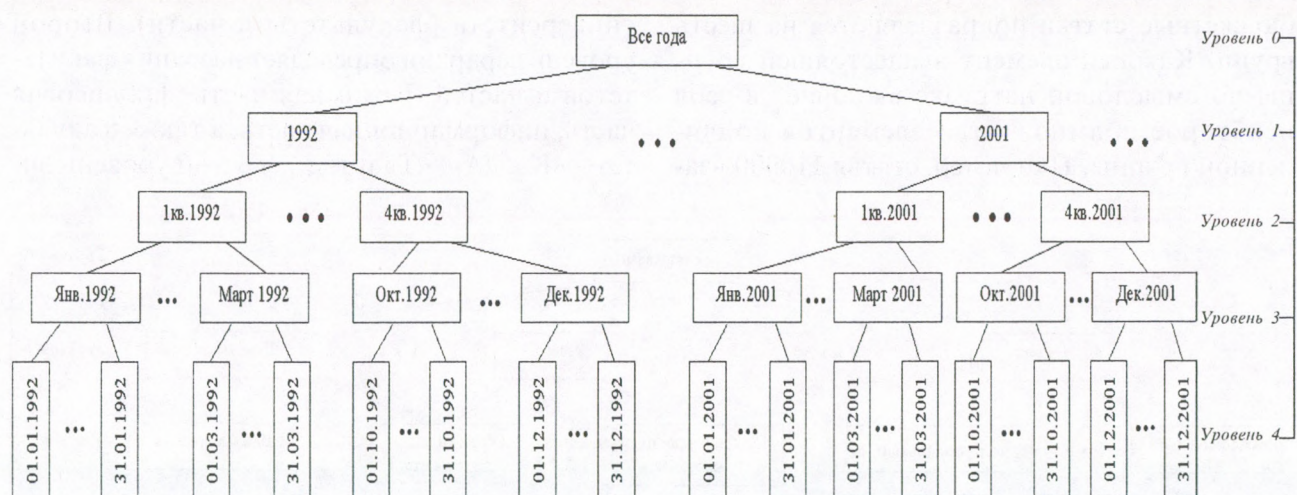


Рис. 5. Иерархия измерения «время»



Рис. 6. Иерархия измерения «подразделения»



ределяет номера структурных подразделений университета (002, 003, 008, 022, 607, 609, 910 и т. д.). Аналогично измерению «источники», четвертый уровень данной иерархии определяет номера внутренних счетов университета (80-4-609-154, 80-8-910-011 и т. п.).

Для большей наглядности приведем пример формирования *среза данных* вышеописанного куба. Представим в виде сводной таблицы динамику расхода финансовых средств по источнику «Министерство науки» за 2000 год в разрезе по внутренним счетам и кварталам (табл. 1). Количество показателей и суммы уменьшены для упрощения демонстрации характерных особенностей предметной области. При формировании описываемого среза данных в первую очередь задаются условия для каждого измерения. Для измерения «внутренние статьи» фиксируется значение «расход». Для измерения «источники» фиксируется значение «Министерство науки» и выбирается уровень детализации «внутренние счета». Для измерения «время» фиксируется значение «2000 г.» и выбирается уровень детализации «кварталы». Для двух оставшихся измерений (подразделения, бюджетные статьи) выбирается нулевой уровень детализации. Затем на основе ячеек куба данных, которые удовлетворяют описываемым условиям, производится агрегирование финансовых показателей в соответствии с заданными для каждого измерения уровнями детализации.

В качестве основы для вышеописанного многомерного представления финансовой информации и последующего использования средств OLAP необходимо построить витрину данных, которая в рамках разрабатываемой модели данных воспринимается как набор отношений различной степени, возможно изменяющихся со временем. При этом витрину данных можно интерпретировать как реляционную БД, отношения которой обладают определен-

ными особенностями. Для описания структуры рассмотренных финансовых данных и представления их в БД потребуется определить нижеперечисленные домены [2, с. 85]. Необходимо отметить, что некоторые из рассматриваемых доменов содержат достаточно большое количество элементов (например, более 3,5 тыс.). В целях экономии мы приводим лишь характерные элементы каждого домена.

Домен *bd* — множество допустимых значений форм финансирования различных видов деятельности, $bd = \{\text{'бюджет', 'договор', 'накладные расходы', 'рефинансирование', 'целевые средства'}\}$.

Домен *vedom* — множество допустимых значений названий министерств и ведомств, а также видов деятельности, в результате которых университет получает финансирование, $vedom = \{\text{'мин. науки', 'мин. образования', 'РФФИ', ..., 'научно-исслед. деятел.', 'образоват. деятел.'}\}$.

Домен *ist* — множество допустимых значений источников финансирования, $ist = \{\text{'4', 'e', 'u', 'm', 'ю', 'a', 'l', '2', ..., '8', 'ц', 'я'}\}$.

Домен *tema* — множество допустимых значений номеров внутренних счетов университета, $tema = \{\text{'80-4-607-153', ..., '80-4-607-171', '80-4-609-154', '80-8-910-011'}\}$.

Домен *chaf* — множество допустимых значений типов структурных подразделений, $chaf = \{\text{'факультеты', 'части'}\}$.

Домен *fakul* — множество допустимых значений названий факультетов и частей, $fakul = \{\text{'А', 'Б', 'Г', 'К', 'Т', 'Ф', ..., 'финансовая часть', 'информационная часть', 'режимная часть'}\}$.

Домен *podr* — множество допустимых значений номеров структурных подразделений, $podr = \{\text{'002', '003', '008', '022', ..., '607', '609', '910'}\}$.

Домен *vnukl* — множество допустимых значений типов финансовых операций, $vnukl = \{\text{'приход', 'расход'}\}$.

Таблица 1

Пример формирования среза данных

Номер внутреннего счета	1-й кв. 2000	2-й кв. 2000	3-й кв. 2000	4-й кв. 2000	2000
80-4-607-153	15.00	17.00	12.00	0.00	44.00
80-4-607-171	18.00	21.00	0.00	14.00	53.00
80-4-609-154	0.00	0.00	28.00	0.00	28.00
<i>Всего</i>					
по «Министерству науки»	33.00	38.00	40.00	14.00	125.00

Домен cod — множество допустимых значений номеров статей внутренней классификации расходов и доходов, $cod = \{ '001', '003', \dots, '100', '120', '305', '801' \}$.

Для адекватного представления системы бюджетных статей расходов определены следующие шесть доменов, которые являются множествами допустимых значений номеров статей бюджетной классификации для каждой группы соответственно:

$bkl1 = \{ '800000', '000000' \};$

$bkl2 = \{ '100000', '200000', '300000' \};$

$bkl3 = \{ '110000', '120000', '130000', \dots, '250000', '270000', '380000' \};$

$bkl4 = \{ '110100', '110200', '110300', \dots, '270200', '380100' \};$

$bkl5 = \{ '110110', '110120', '110130', \dots, '110140', '380240', '380410' \};$

$cod2 = \{ '110111', '110112', \dots, '120125', '120126' \}.$

Поскольку измерение «время» в рамках вышеописанного многомерного представления финансовых данных представлено следующими уровнями детализации: отдельные даты, месяцы, кварталы, годы, для представления каждого уровня определены домены, устанавливающие соответствующие множества допустимых значений:

$god = \{ '1992', '1993', '1994', \dots, '1999', '2000', '2001' \};$

$kvert = \{ '1кв.1992', '2кв.1992', \dots, '2кв.2001', '3кв.2001', '4кв.2001' \};$

$mes = \{ 'январ.1992', 'февр.1992', \dots, 'окт.2001', 'нояб.2001', 'дек.2001' \};$

$chisl = \{ '01.01.92', '02.01.92', \dots, '29.12.01', '30.12.01', '31.12.01' \}.$

Домен $dsum$ — множество допустимых значений сумм расходов и поступлений финансовых средств, $dsum = \{ m/100 : m \in Z \}$, где Z — множество целых чисел.

Согласно реляционной модели данных [4, с. 288], будем использовать следующие определения и обозначения. **Отношение** на множествах D_1, \dots, D_n формально определяется как подмножество множества расширенного декартова произведения χ n -множеств D_1, \dots, D_n , которое в свою очередь формально определяется выражением $\chi(D_1, \dots, D_n) = \{ (d_1, \dots, d_n) : d_j \in D_j, j=1, \dots, n \}$. Элементы отношения представляют собой n -ки (упорядоченные цепочки символов, также называемые кортежами), содержащие n элементов, где n — степень отношения. Множества D_1, \dots, D_n называются доме-

нами отношения. Каждое вхождение домена в декартово произведение называется атрибутом. Схему отношения будем задавать строкой $R(D_1, D_2, \dots, D_n)$, где R — имя отношения; n — степень отношения; D_i — имя i -го домена. При этом введем условные обозначения для ключей отношения. Домены, образующие первичный ключ, будем выделять жирным шрифтом, а домены, являющиеся внешним ключом отношения, будем выделять подчеркиванием. Отметим, что рассматриваемые отношения являются нормализованными домено-неупорядоченными отношениями [8, с. 380]. Для целей создания витрины данных на основе описываемой структуры данных все отношения разделим на две группы. К первой из них относятся отношения, определяющие измерения, на основе которых проводится анализ финансовых данных и взаимосвязь уровней иерархии соответствующих измерений. Ко второй группе относятся отношения, определяющие собственно анализируемые финансовые данные.

Первая группа

Отношение S_Ist — определяет измерение «источники»,

$S_Ist(tema, ist, vedom, bd).$

Отношение S_Vncd — определяет измерение «внутренние статьи»,

$S_Vncd(cod, vnukl).$

Отношение S_Time — определяет измерение «время»,

$S_Time(chisl, mes, kvart, god).$

Отношение S_Podr — определяет измерение «подразделения»,

$S_Podr(tema, podr, fakul, chaf).$

Отношение S_Spisok — определяет измерение «бюджетные статьи»,

$S_Spisok(cod2, bkl5, bkl4, bkl3, bkl2, bkl1).$

Вторая группа

Отношение M_Olap — определяет суммы расходов и поступлений финансовых средств и взаимосвязь анализируемых данных с измерениями,

$M_Olap(tema, chisl, cod, cod2, dsum).$

Для большей наглядности перечислим все вышеназванные домены, отношения и приведем диаграмму витрины данных [7], на которой схематично отразим отношения и связи между ними (рис. 7). Для отображения первичных и внешних ключей отношения будем использовать ранее введенные обозначения.



Специфика предметной области и цели построения структуры данных таковы, что отношения первой группы обладают особенностями, для описания которых введем дополнительные обозначения. Обозначим через M_R — множество доменов отношения R , например $M_{S_Vncd} = \{cod, vnukl\}$. Проекцию (в смысле реляционной алгебры [8, с. 383]) отношения R по домену D обозначим как $\pi_D(R)$, т. е. $\pi_D(R)$ — это унарное отношение, по существу являющееся множеством содержащихся в некоторый момент времени в отношении R элементов домена D . Обозначим через Dim — множество отношений первой группы, $Dim = \{S_Ist, S_Vncd, S_Time, S_Podr, S_Spisok\}$. Используя введенные обозначения, получаем: в каждый момент времени для любых отношения R и домена D , таких, что R принадлежит первой группе и D является доменом отношения R , справедливо то, что проекция $\pi_D(R)$ отношения R по домену D содержит все элементы D , что формально записывается:

$$\forall R, D : R \in Dim \ \& \ D \in M_R \rightarrow \pi_D(R) = D. \quad (1)$$

Однако соотношение (1) не справедливо для отношений второй группы. Про M_Olap можно утверждать лишь следующее: в каждый момент времени для любого домена D такого, что D является доменом отношения M_Olap справедливо то, что проекция $\pi_D(M_Olap)$ отношения M_Olap по домену D является подмножеством домена D , что формально записывается как

$$\forall D : D \in M_{M_Olap} \rightarrow \pi_D(M_Olap) \subset D.$$

Для обозначения функциональной зависимости [4, с. 303] примем, что атрибут b отношения R функционально зависит от атрибута a отношения R , если в каждый момент времени каждое значение в a не имеет более одного значения в b , соотношенного с ним посредством R . Будем писать $R.a \Rightarrow R.b$, если b функционально зависит от a в отношении R , и $R.a \not\Rightarrow R.b$, если такой зависимости нет.

В отношениях первой группы существуют следующие функциональные зависимости:

$$\begin{aligned} S_Ist.tema &\Rightarrow S_Ist.ist, S_Ist.ist \Rightarrow S_Ist.vedom, \\ S_Ist.vedom &\Rightarrow S_Ist.bd; \\ S_Vncd.cod &\Rightarrow S_Vncd.vnukl; S_Time.chisl \Rightarrow \\ &\Rightarrow S_Time.mes, S_Time.mes \Rightarrow S_Time.kvart, \end{aligned}$$

$$\begin{aligned} S_Time.kvart &\Rightarrow S_Time.god; \\ S_Podr.tema &\Rightarrow S_Podr.podr, S_Podr.podr \Rightarrow \\ &\Rightarrow S_Podr.fakul, S_Podr.fakul \Rightarrow S_Podr.chaf; \\ S_Spisok.cod2 &\Rightarrow S_Spisok.bkl5, S_Spisok.bkl5 \Rightarrow \\ &\Rightarrow S_Spisok.bkl4, S_Spisok.bkl4 \Rightarrow \\ &\Rightarrow S_Spisok.bkl3; \\ S_Spisok.bkl3 &\Rightarrow S_Spisok.bkl2, S_Spisok.bkl2 \Rightarrow \\ &\Rightarrow S_Spisok.bkl1. \end{aligned}$$

При этом, если представить любую из вышеперечисленных функциональных зависимостей в общем виде, т. е. $R.a \Rightarrow R.b$, то, учитывая особенности предметной области, во всех случаях справедливо, что $R.b \not\Rightarrow R.a$.

Рассмотренные функциональные зависимости демонстрируют, что в отношениях первой группы существуют транзитивные зависимости. В результате чего получаем, что вышеописанные отношения находятся во второй нормальной форме, т. к. во всех отношениях атрибуты, не входящие в состав первичного ключа, полностью зависят от последнего [4, с. 304–309]. Конечно, можно данные отношения привести к третьей нормальной, однако это не делается по соображениям производительности (подробнее данный вопрос обсуждается в последующих главах).

2. Синтаксис языка манипулирования данными

Для проверки адекватности проектируемой витрины данных вышеописанному многомерному представлению финансовой информации представим язык манипулирования данными. Поскольку при построении витрины данных использовалась реляционная модель данных, то при построении языка будем стараться, по возможности, придерживаться канонического реляционного исчисления [2, 4, 8, 9]. Таким образом, основой применяемой конструкции будет

$$\text{целевой список} \mid \text{квалифицирующее выражение}. \quad (2)$$

В различных источниках [2, с. 191; 4, с. 296] вертикальная черта иногда заменяется двоеточием или словом «где». Вместе с тем для сохранения возможностей построения витрин данных в язык будут добавлены средства указания на измерения.

Введем алфавит языка, который в ходе изложения будет дополнен: $\alpha_1, \alpha_2, \alpha_3, \dots$ — ска-

Домены

bd = {'бюджет', 'договор', 'накладные расходы', 'рефинансирование', 'целевые средства'}
vedom = {'мин. науки', 'мин. образования', 'РФФИ', ... , 'научно-исслед. деятел.', 'образоват. деятел.'}
ist = {'4', 'е', 'и', 'т', 'ю', 'а', '1', '2', ... , '8', 'ц', 'я'}
tema = {'80-4-607-153', ... , '80-4-607-171', '80-4-609-154', '80-8-910-011'}

chaf = {'факультеты', 'части'}
fakul = {'А', 'Б', 'Т', 'К', 'Т', 'Ф', ... , 'финансовая часть', 'информационная часть', 'режимная часть'}
podr = {'002', '003', '008', '022', ... , '607', '609', '910'}

vnukl = {'приход', 'расход'}
cod = {'001', '003', ... , '100', '120', '305', '801'}

bkl1 = {'800000', '000000'}
bkl2 = {'100000', '200000', '300000'}
bkl3 = {'110000', '120000', '130000', ... , '250000', '270000', '380000'}
bkl4 = {'110100', '110200', '110300', ... , '270200', '380100'}
bkl5 = {'110110', '110120', '110130', ... , '110140', '380240', '380410'}
cod2 = {'110111', '110112', ... , '120125', '120126'}

god = {'1992', '1993', '1994', ... , '1999', '2000', '2001'}
kvart = {'1кв.1992', '2кв.1992', ... , '2кв.2001', '3кв.2001', '4кв.2001'}
mes = {'январ.1992', 'фев.1992', ... , 'окт.2001', 'нояб.2001', 'дек.2001'}
chisl = {'01.01.92', '02.01.92', ... , '29.12.01', '30.12.01', '31.12.01'}

dsum = {*m*/100 : *mOZ* }, где *Z* — множество целых чисел

Отношения

Первая группа (измерения)

Вторая группа (финансовые показатели)

S_Ist (*tema*, *ist*, *vedom*, *bd*).

M_Olap (*tema*, *chisl*, *cod*, *cod2*, *dsum*).

S_Vncd (*cod*, *vnukl*).

S_Time (*chisl*, *mes*, *kvart*, *god*).

S_Podr (*tema*, *podr*, *fakul*, *chaf*).

S_Spiskok (*cod2*, *bkl5*, *bkl4*, *bkl3*, *bkl2*, *bkl1*).

Диаграмма

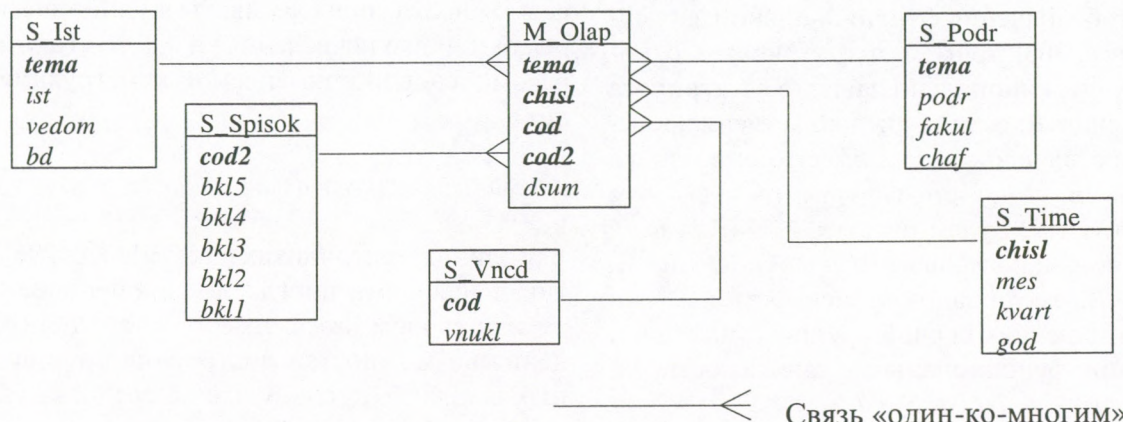


Рис. 7. Проектируемая витрина данных



лярные константы; *rel*, *var*, *attr* (возможно с индексами) — идентификаторы, представляющие имя отношения, имя переменной выборки, имя атрибута соответственно; \exists , *and*, *or*, *not* — логические символы; (,) — разделители. Условимся, что любая конструкция языка, заключенная в квадратные скобки «[]», может быть опущена.

Область значений, по которой пробегает переменная выборки, определим как

$$var \in rel, \quad (3)$$

где *var* — переменная выборки, *rel* — имя отношения.

Для определения значения отдельного компонента (соответствующего конкретному атрибуту отношения) переменной выборки введем конструкцию, которую будем называть **уточненное имя атрибута**:

$$var.attr, \quad (4)$$

где *attr* — имя атрибута отношения, на котором определена переменная выборки *var*.

Введем понятие **элементарной формулы**, которую определим как

$$comp_1 q comp_2, \quad (5)$$

где *q* — скалярный оператор сравнения ($q = \{=, \neq, >, <, \geq, \leq\}$); каждый операнд (*comp₁* и *comp₂*) является либо скалярной константой *a*, либо уточненным именем атрибута (4). Разрешаются любые сочетания: «*var₁.attr q var₂.attr*», «*var.attr q a*», «*a q var.attr*», «*a₁ θ a₂*».

Правильно построенную формулу (ППФ) зададим следующим образом:

1) каждая элементарная формула (5) считается ППФ;

2) если *f* — ППФ, то и *not f* — тоже ППФ;

3) если *f₁* — ППФ и *f₂* — ППФ, то *f₁ and f₂* и *f₁ or f₂* — ППФ;

4) если *f* — ППФ, в которой *var* используется как свободная переменная, то $\exists var(f)$ — ППФ, а переменная *var* является связанной в этой формуле.

Приведем правила, определяющие, каким будет экземпляр переменной выборки — свободным или связанным [2, с.193]. Под экземпляром переменной выборки в ППФ будем понимать наличие имени переменной в ППФ в

виде конструкции (4) или наличие имени переменной, непосредственно следующего за логическим символом \exists .

Правила:

— в элементарной формуле все экземпляры переменных выборки (если такие есть) являются свободными;

— экземпляры переменных выборки в ППФ вида «*not f*» свободны или связаны в зависимости от того, свободны или связаны они в *f*;

— экземпляры переменных выборки в ППФ вида «*f₁ and f₂*» и «*f₁ or f₂*» свободны или связаны в зависимости от того, свободны или связаны они в *f₁* и *f₂*;

— экземпляры переменной выборки *var*, которые свободны в ППФ *f*, связаны в ППФ вида « $\exists var(f)$ ». Экземпляры других переменных выборок в такой формуле свободны или связаны в зависимости от того, свободны или связаны они в *f*.

В соответствии с (2) выражение *exp* определим следующей конструкцией:

$$(tic) [WHERE formula], \quad (6)$$

где *tic* — целевой список, *formula* — квалифицирующее выражение, которое является ППФ. В выражении *exp* (6) каждая свободная переменная в *formula* (если такая есть) должна быть указана в целевом списке элементов *tic*. В свою очередь целевой список *tic* имеет вид *ele₁, ele₂, ..., ele_n*, где $n \geq 1$. Каждый элемент целевого списка *ele_i* ($i = 1, \dots, n$) — конструкция вида *var.attr [AS attr₁]* или конструкция вида $\alpha AS attr_1$, где *var.attr* — уточненное имя атрибута, α — скалярная константа *attr₁* — имя атрибута результата вычисления выражения *exp*. Хотя определение элемента целевого списка *ele* придется дополнить, сделано это будет ниже.

Функцию агрегации SUM определим следующим образом:

$$SUM(exp, attr), \quad (7)$$

где *exp* — выражение вида (6); *attr* — имя атрибута результата вычисления выражения *exp*.

Используя конструкцию (7), элемент целевого списка *ele* также может быть представлен в виде *SUM(exp, attr) AS attr₁*. Отметим, что в данном случае, как и в случае со скалярной константой, часть «*AS attr₁*» является обязательной.

3. Обработка типовых запросов

В рамках многомерной модели данных [3] существуют следующие основные операции манипулирования данными: формирование срезов куба данных, агрегирование, детализация и поворот. Для проверки адекватности проектируемой витрины данных вышеописанному многомерному представлению финансовой информации необходимо рассмотреть процесс формирования результатов специфических операций манипулирования многомерными данными на основе результатов запросов

к витрине данных, которая в рамках настоящей работы представлена реляционной БД. Для упрощения графического отображения многомерного представления результатов запросов и возможности проверки результатов запросов вручную условимся, что куб данных (рис. 8) будет содержать расходы финансовых средств за 2000 год и основываться на трех измерениях (из пяти ранее определенных): «источники», «время» и «внутренние статьи». Необходимо отметить, что данные упрощения никак не повлияют на демонстрацию

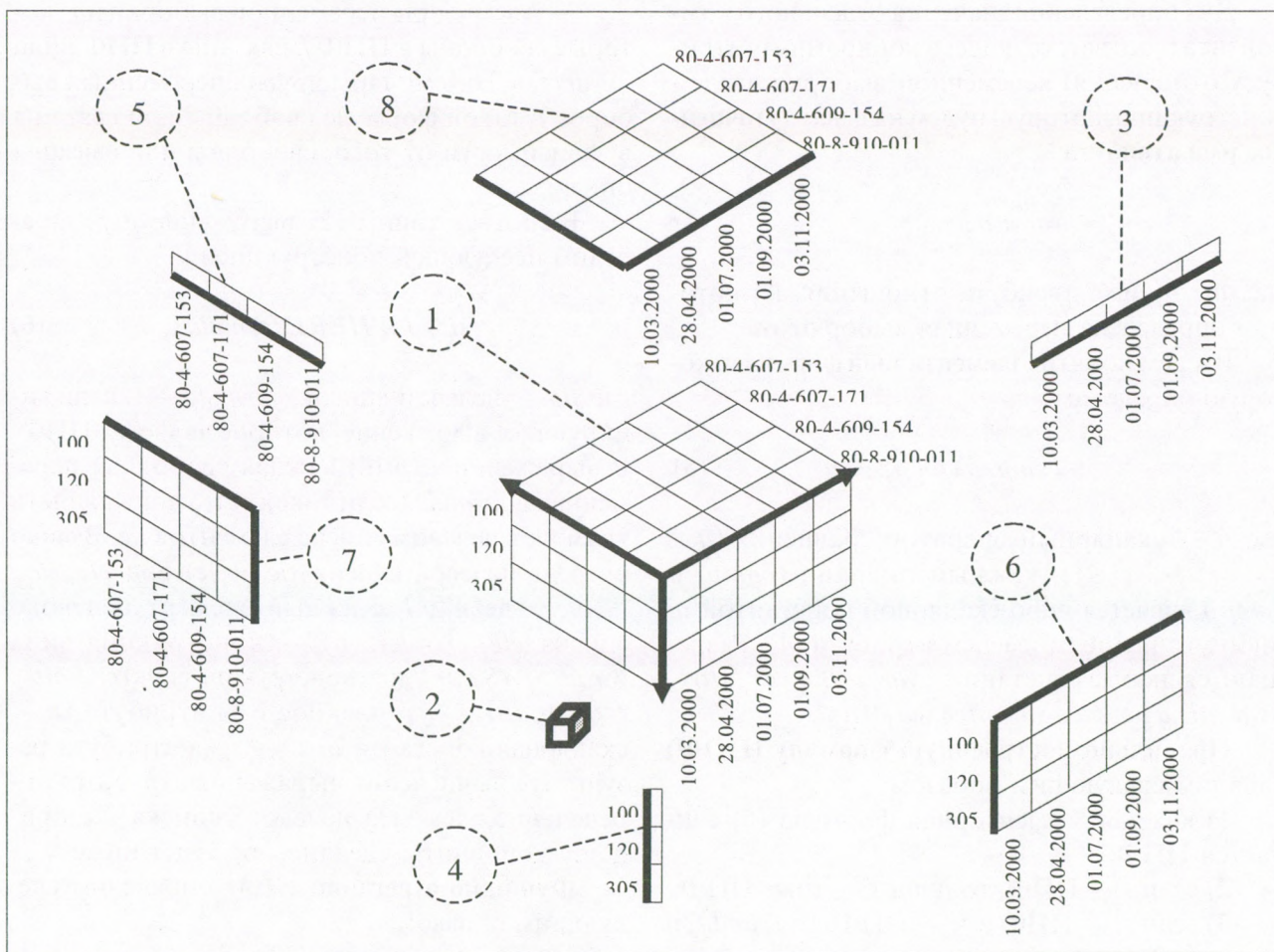


Рис.8. Основные компоненты многомерного представления финансовых данных:

- 1 — куб данных (детализированные данные);
- 2 — агрегированный показатель (суммирование расходов по всем измерениям);
- 3 — срез данных (фиксация измерения «время», суммирование по остальным измерениям);
- 4 — срез данных (зафиксированное измерение «внутренние статьи»);
- 5 — срез данных (фиксация измерения «источники»);
- 6 — срез данных (зафиксированные измерения «время», «внутренние статьи»);
- 7 — срез данных (фиксация измерений «источники», «внутренние статьи»);
- 8 — срез данных (зафиксированные измерения «время», «источники»).



характерных особенностей обработки типовых за-просов.

В соответствии с этим витрина данных (рис. 9) будет содержать четыре отношения из вышеописанных шести (для первичных и внешних ключей используются прежние обозначения):

$S_Ist(tema, ist, vedom, bd);$

$S_Time(chisl, mes, kvart, god);$
 $S_Vncd(cod, vnukl);$
 $M_Olap(tema, chisl, cod, dsum).$

Далее подразумевается, что заданы переменные выборки i, t, v, x , что формально запишем как
 $i_i \in S_Ist; t_i \in S_Time; v_i \in S_Vncd; x_i \in M_Olap.$

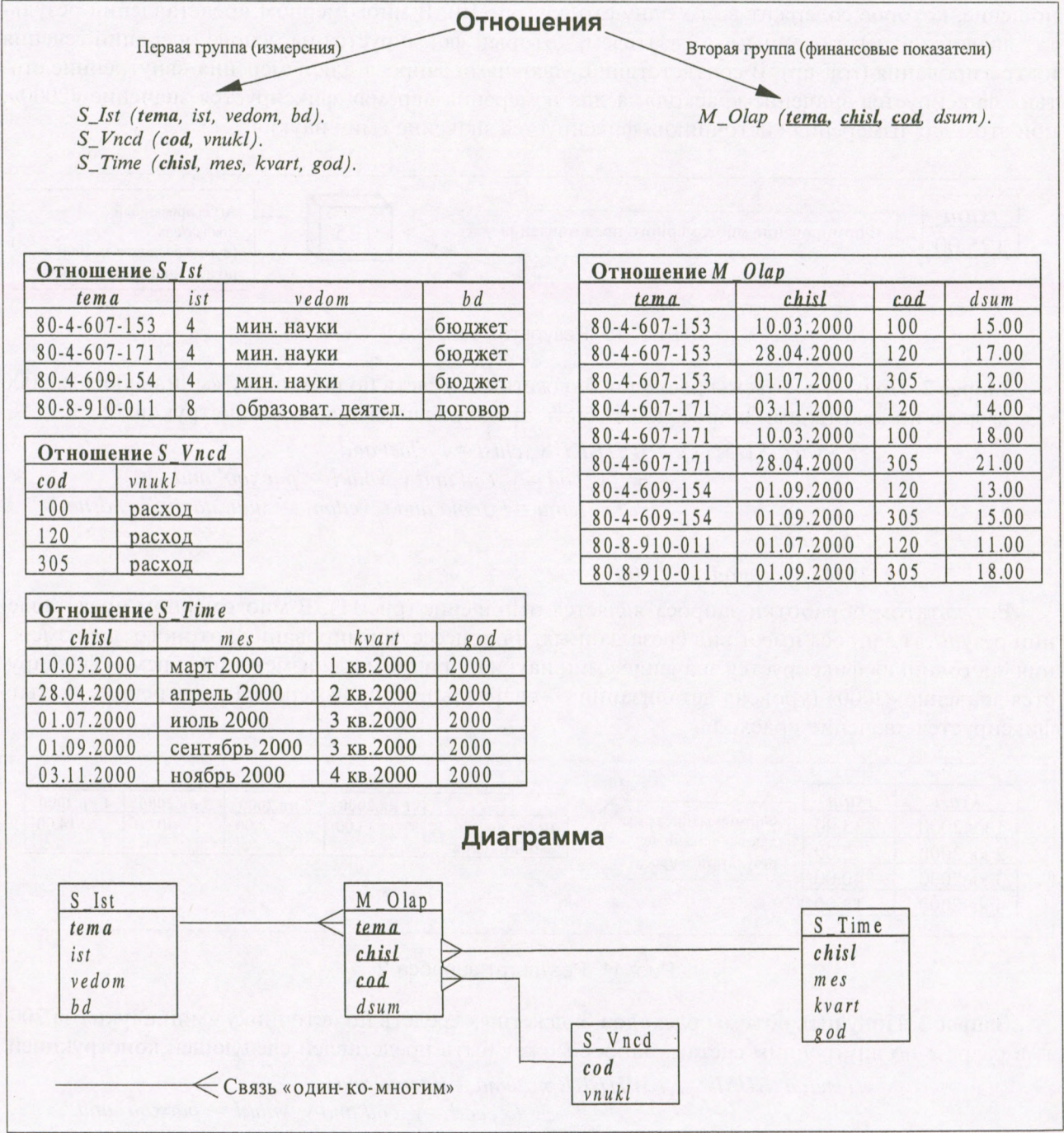


Рис. 9. Используемая в примерах витрина данных

Используя вышеописанный язык, приведем примеры типовых запросов к витрине данных и сравним результаты их обработки с результатами многомерных операций.

Запрос 1. Получить объем расходов бюджетных средств по источнику «мин.науки» за 2000 г. Данный запрос может быть представлен следующей конструкцией:

$$(SUM((x_i) \text{ WHERE } \exists i_i(x_i.tema = i_i.tema \text{ and } i_i.vedom = 'мин. науки' \text{ and } \\ \exists v_i(x_i.cod = v_i.cod \text{ and } v_i.vnukl = 'расход' \text{ and } \\ \exists t_i(x_i.chisl = t_i.chisl \text{ and } t_i.god = '2000'))), dsum) \\ AS rsum)$$

Результатом обработки запроса, в соответствии с целевым списком, является унарное отношение, которое содержит всего одну строку (рис.10). В многомерном представлении результат является агрегированным показателем, который формируется на основе операций сечения и агрегирования (roll-up). В соответствии с условиями запроса для измерения «внутренние статьи» фиксируется значение «расход», а для измерения «время» фиксируется значение «2000», при этом для измерения «источники» фиксируется значение «мин.науки».

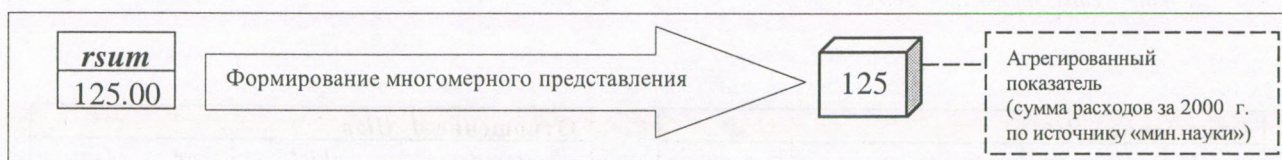


Рис. 10. Результат запроса 1

Запрос 2. Получить объемы расходов бюджетных средств по источнику «мин.науки» за 2000 г. в разрезе по кварталам. Запрос может быть представлен следующей конструкцией:

$$(t_i.kvart, SUM((x_i) \text{ WHERE } x_i.chisl = t_i.chisl \text{ and } \\ \exists v_i(x_i.cod = v_i.cod \text{ and } v_i.vnukl = 'расход' \text{ and } \\ \exists i_i(x_i.tema = i_i.tema \text{ and } i_i.vedom = 'мин. науки'))), dsum) \\ AS rsum) \\ \text{WHERE } t_i.god = '2000'$$

Результатом обработки запроса является отношение (рис.11). В многомерном представлении результат запроса имеет вид среза данных, в процессе формирования которого для измерения «источники» фиксируется значение «мин.науки», при этом для измерения «время» фиксируется значение «2000» (уровень детализации — кварталы), а для измерения «внутренние статьи» фиксируется значение «расход».

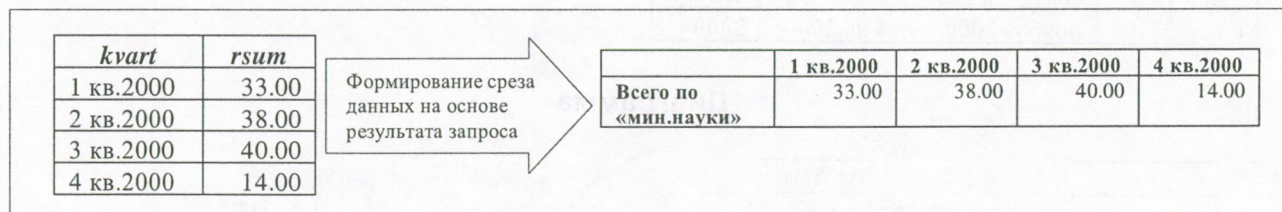


Рис. 11. Результат запроса 2

Запрос 3. Получить объемы расходов бюджетных средств по источнику «мин.науки» за 2000 г. в разрезе по внутренним счетам. Запрос может быть представлен следующей конструкцией:

$$(i_i.tema, SUM((x_i) \text{ WHERE } x_i.tema = i_i.tema \text{ and } \\ \exists v_i(x_i.cod = v_i.cod \text{ and } v_i.vnukl = 'расход' \text{ and } \\ \exists t_i(x_i.chisl = t_i.chisl \text{ and } t_i.god = '2000'))), dsum) \\ AS rsum) \\ \text{WHERE } i_i.vedom = 'мин. науки'$$



Результатом обработки запроса является отношение (рис. 12). В многомерном представлении результат запроса имеет вид среза данных, в процессе формирования которого для измерения «источники» фиксируется значение «мин. науки» (уровень детализации — внутренние счета), при этом для измерения «время» фиксируется значение «2000», а для измерения «внутренние статьи» фиксируется значение «расход».

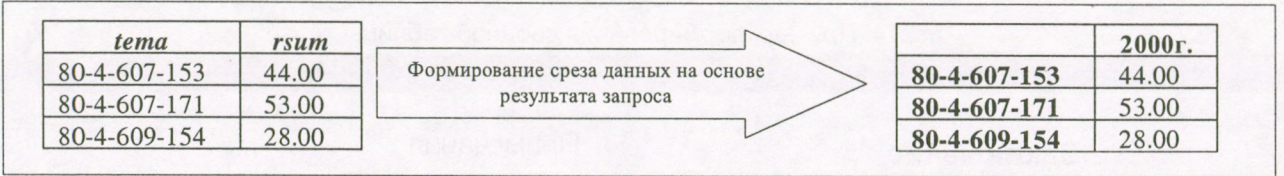


Рис. 12. Результат запроса 3

Запрос 4. Получить объемы расходов бюджетных средств по источнику «мин. науки» за 2000 г. в разрезе по внутренним счетам и кварталам. Запрос может быть представлен следующей конструкцией:

```
(ir.tema, tr.kvart, SUM((xr) WHERE xr.tema = ir.tema and xr.chisl = tr.chisl and  
∃vr(xr.cod = vr.cod and vr.vnukl = 'расход'), dsum)  
AS rsum)  
WHERE ir.vedom = 'мин. науки' and tr.god = '2000'
```

Результатом обработки запроса является отношение (рис. 13). В многомерном представлении результат запроса имеет вид среза данных, в процессе формирования которого для измерения «источники» фиксируется значение «мин.науки» (уровень детализации — внутренние счета), при этом для измерения «время» фиксируется значение «2000» (уровень детализации — кварталы), для измерения «внутренние статьи» фиксируется значение «расход».

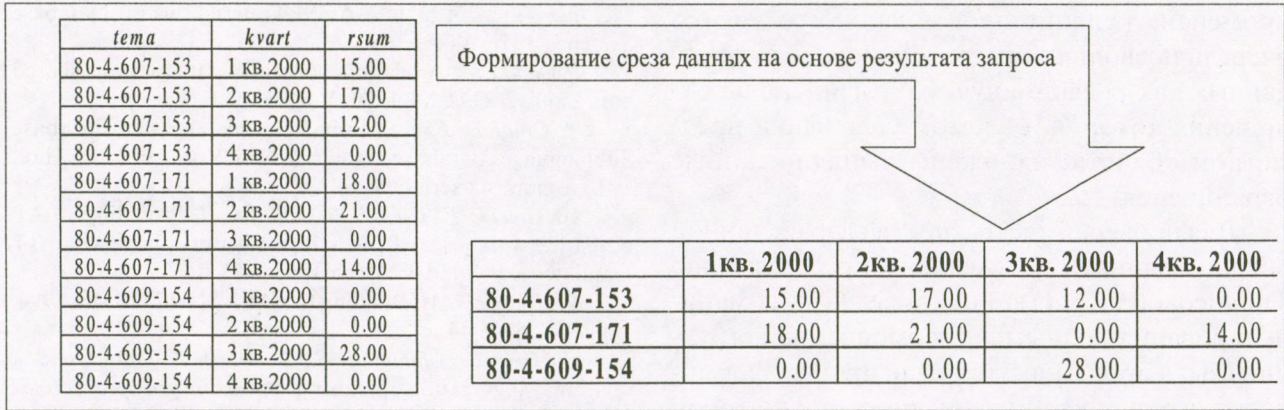


Рис. 13. Результат запроса 4

Как это видно из вышерассмотренных примеров, результаты запросов 1–4 (срезы данных и агрегированные показатели) могут быть скомбинированы и представлены в виде сводной таблицы (рис. 14), которая отображает динамику расхода бюджетных средств по источнику «мин. науки» за 2000г., т. е. является полным аналогом сводной таблицы, приведенной в качестве примера в начале работы (см. табл. 1). Таким образом, данный пример демонстрирует полное соответствие между многомерным представлением финансовых данных и разрабатываемой витриной данных.

		1 кв. 2000	2 кв. 2000	3 кв. 2000	4 кв. 2000	2000 г.	
Результат запроса 4	80-4-607-153	15.00	17.00	12.00	0.00	44.00	Результат запроса 3
	80-4-607-171	18.00	21.00	0.00	14.00	53.00	
	80-4-609-154	0.00	0.00	28.00	0.00	28.00	
Результат запроса 2	Всего по «мин. науки»	33.00	38.00	40.00	14.00	125.00	Результат запроса 1

Рис. 14. Пример формирования сводной таблицы

Заключение

Предложенная модель данных обеспечивает возможность построения витрины данных, которая в свою очередь используется в качестве основы для многомерного представления финансовой информации и последующего анализа ее с помощью инструментальных средств оперативной аналитической обработки (OLAP).

По результатам анализа предметной области (управление финансовыми данными) разработана структура многомерного представления финансовой информации в виде куба данных (определены анализируемые данные и базовые измерения, на множестве значений которых установлены отношения иерархии).

При построении витрины данных, используемой в качестве основы для куба данных, применена реляционная модель, что в свою очередь позволило интерпретировать витрину данных как реляционную базу данных, на отношения которой, с учетом специфики предметной области, наложены вышеописанные ограничения.

Для проверки адекватности разработанной витрины данных многомерному представлению финансовой информации приведен язык манипулирования данными. С помощью конструкций языка продемонстрирован процесс формирования результатов специфических многомерных операций манипулирования данными (детализация, агрегирование, формирование срезов) на основе запросов к витрине данных.

Примечания

1. Горелов Б. А., Рыжухина В. А. Корпоративная система — настоящее и будущее // МИФИ'98: Материалы научной сессии. М., 1998. Т. 5. С. 174.
2. Дейт К. Дж. Введение в системы баз данных. 6-е изд. / Пер. с англ. Киев; М.; СПб.: Изд. дом «Вильямс», 1999. 848 с.
3. Козаловский М. Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002. 800 с.
4. Кузин Л. Т. Основы кибернетики: В 2 т. Т. 2: Основы кибернетических моделей. М.: Энергия, 1979. 584 с.
5. Левенец И. А. Методы и средства автоматизации проектирования и эксплуатации хранилищ данных (в сфере электроэнергетики и регионального управления): Автореф. дис. ... канд. техн. наук: 05.13.12. Иваново: ИГЭУ, 2001.
6. Щавелев Л. В. Автоматизация проектирования систем оперативной аналитической обработки данных (на примере информационно-аналитических систем в энергетике): Автореф. дис. ... канд. техн. наук: 05.13.12, 05.13.14. Иваново: ИГЭУ, 1999.
7. Chen Peter P. The Entity-Relationship Model Toward a Unified View of Data // ACM Transactions on Database Systems. 1976. Vol. 1. № 1.
8. Codd E. F. A relational model of data for large shared data banks // CACM. 1970. Vol. 13, № 6.
9. Codd E. F. Relational completeness of database Sublanguages database systems // New York: Prentice Hall. 1972 (Symposia series. № 6).
10. Codd E. F., Codd S. B., Salley C. T. Providing OLAP (On-Line Analytical Processing) to user-analysts: an IT mandate. E.F.Codd & Associates, 1993.
11. Demarest M. Building the Data Mart // DBMS. 1994. Vol. 7, № 8. P. 44.
12. Gray J., Chaudhuri S., Bosworth A. Data Cube: a relational aggregation operator generalizing group-by, cross-tab, and sub-totals // Data Mining and Knowledge Discovery. 1997. № 1. P. 29–53.
13. Mumick I. S., Quass D., Mumick B. S. Maintenance of Data Cubes and Summary Tables in a Warehouse // Stanford University, Database Group, 1996.
14. Raden N. Star Schema. Santa Barbara, CA: Archer Decision Sciences, Inc., 1995–1996. <http://members.aol.com/nraden/str101.htm>